DOCUMENT RESUME

ED 440 846 SE 063 430

AUTHOR	Steen, Lynn Arthur
TITLE	Assessing Assessment.
PUB DATE	1999-00-00
NOTE	10p.; In: Assessment Practices in Undergraduate Mathematics. Washington, DC: Mathematical Association of America, 1999.
AVAILABLE FROM	For full text: http://www.stolaf.edu/people/steen/papers/assessment.html.
PUB TYPE	Guides - Non-Classroom (055)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*College Mathematics; *Educational Assessment; Elementary Secondary Education; *Evaluation Methods; Higher Education; *Mathematics Education

ABSTRACT

Assessment not only places value, it also identifies which elements to value. In this era of accountability, the constituents of educational assessment are not just students, faculty, and administrators, but also parents, legislators, journalists, and the public. For these broader audiences, simple numerical indicators of student performance take on totemic significance. The spotlight of educational improvement often falls first and brightest on mathematics. This paper first discusses college mathematics and principles of assessment, then elaborates on the implementations of assessment. (Contains 21 references.) (ASK)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 440 846

Assessing Assessment

by
Lynn Arthur Steen

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

L. Steen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

D Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

SE060430

Assessing Assessment

Lynn Arthur Steen, St. Olaf College

In *Assessment Practices in Undergraduate Mathematics.* Bonnie Gold, *et al.*, Editors. Washington DC: Mathematical Association of America, 1999.

(The Preface to an MAA volume that contains several dozen reports from different campuses of diverse assessment activities in the mathematical sciences.)

We open letters from the city assessor with trepidation since we expect to learn that our taxes are about to go up. Mathematicians typically view academic assessment with similar emotion. Some react with indifference and apathy, others with suspicion and hostility. Virtually no one greets a request for assessment with enthusiasm. Assessment, it often seems, is the academic equivalent of death and taxes: an unavoidable waste.

In ancient times, an assessor (from *ad + sedere)* was one who sat beside the sovereign to provide technical advice on the value of things that were to be taxed. Only tax collectors welcomed assessors. Tradition, self-interest, and common sense compel faculty to resist assessment for many of the same reasons that citizens resist taxes.

Yet academic sovereigns (read: administrators) insist on assessment. For many reasons, both wise and foolish, administrators feel compelled to determine the value of things. Are students learning what they should? Do they receive the education they have been promised? Do our institutions serve well the needs of all students? Are parents and the public receiving value for their investment in education? Are educational programs well suited to the needs of students? Do program benefits justify costs? Academic sovereigns ask these questions not to impose taxes but to determine institutional priorities and allocate future resources.

What we assess defines what we value [Wiggins, 1990]. Students' irreverent questions ("Will it be on the test?") signal their understanding of this basic truth. They know, for example, that faculty who assess only calculation do not really value understanding. In this respect, mathematics faculty are not unlike their students: while giving lip service to higher goals, both faculty and students are generally satisfied with evidence of routine performance. Mathematics departments commonly claim to want their majors to be capable of solving real-world problems and communicating mathematically. Yet these goals ring hollow unless students are evaluated by their ability to identify and analyze problems in real-world settings and communicate their conclusions to a variety of audiences. Assessment not only places value on things, but also identifies the things we value.

In this era of accountability, the constituencies of educational assessment are not just students, faculty, and administrators, but also parents, legislators, journalists, and the public. For these broader audiences, simple numerical indicators of student performance take on totemic significance. Test acronyms (SAT, TIMSS, NAEP, AP, ACT, GRE) compete with academic subjects (mathematics, science, history) as the public vocabulary of educational discourse. Never mind that GPA is more a measure of student compliance than of useful knowledge, or that SAT scores reflect relatively narrow test-taking abilities. These national assessments have become, in the public eye, surrogate definitions of education. In today's assessment-saturated environment, mathematics *is* the mathematics that is

ED 434 343

tested.

College Mathematics

In most colleges and universities, mathematics is arguably the most critical academic program. Since students in a large majority of degree programs and majors are required to take (or test out of) courses in the mathematical sciences, on most campuses mathematics enrollments are among the highest of any subject. Yet for many reasons, the withdrawal and failure rates in mathematics courses are higher than in most other courses. The combination of large enrollments and high failure rates makes mathematics departments responsible for more student frustration–and drop-out–than any other single department.

What's more, in most colleges and universities mathematics is the most elementary academic program. Despite mathematics' reputation as an advanced and esoteric subject, the average mathematics course offered by most postsecondary institutions is at the high-school level. Traditional postsecondary level mathematics--calculus and above--accounts for less than 30% of the 3.3 million mathematical science enrollments in American higher education [Loftsgaarden, 1997].

Finally, in most colleges and universities, mathematics is the program that serves the most diverse student needs. In addition to satisfying ordinary obligations of providing courses for general education and for mathematics majors, departments of mathematical sciences are also responsible for developmental courses for students with weak mathematics backgrounds; for service courses for programs ranging from agriculture to engineering and from business to biochemistry; for the mathematical preparation of prospective teachers in elementary, middle, and secondary schools; for research experiences to prepare interested students for graduate programs in the mathematical sciences; and, in smaller institutions, for courses and majors in statistics, computer science, and operations research.

Thus the spotlight of educational improvement often falls first and brightest on mathematics. In the last ten years alone, new expectations have been advanced for school mathematics [NCTM, 1989], for college mathematics below calculus [AMATYC, 1995], for calculus [Douglas, 1986; Steen, 1988; Roberts, 1996], for statistics [Hoaglin & Moore, 1992], for undergraduate mathematics [Steen, 1989], for departmental goals [MSEB, 1991] and for faculty rewards [JPBM, 1994]. Collectively, these reports convey new values for mathematics education that focus departments more on student learning than on course coverage; more on student engagement than on faculty presentation; more on broad scholarship than on narrow research; more on context than on techniques; more on communication than on calculation. In short; these reports stress mathematics for all rather than mathematics for the few, or (to adopt the slogan of calculus reform) mathematics as "a pump, not a filter."

Principles of Assessment

Assessment serves many purposes. It is used, among other things, to diagnose student needs, to monitor student progress, to give students grades, to judge teaching effectiveness, to determine raises and promotions, to evaluate curricula and programs, and to decide on allocation of resources. During planning (of courses, programs, curricula, majors) assessment addresses the basic questions of why, who, what, how, and when. In the thick of things (in mid-course or mid-project) so-called formative assessment monitors implementation (is the plan going as expected?) and progress (are students advancing adequately?). At the summative stage–which may be at the end of a class period, or of a course, or of a special project–assessment seeks to record impact (both intended and unintended), to compare outcomes with goals, to rank students, and to stimulate action either to modify, extend, or replicate.

Several years ago a committee of the Mathematical Association of America undertook one of the very first efforts in higher education to comprehend the role of assessment in a single academic discipline [Madison, 1992; CUPM, 1995]. Although this committee focused on assessing the mathematics major, its findings and analyses apply to most forms of assessment. The committee's key finding is that assessment, broadly defined, must be a cyclic process of setting goals, selecting methods, gathering evidence, drawing inferences, taking action, and then re-examining goals and methods. Assessment is the feedback loop of education. As the system of thermostat, furnace, and radiators can heat a house, so a similar assessment system of planning, instruction, and evaluation can help faculty develop and provide effective instructional programs. Thus the first principle: *Assessment is not a single event, but a continuous cycle.*

The assessment cycle begins with goals. If you want heat, then you must measure temperature. On the other hand, if it is humidity that is needed, then a thermostat won't be of much use. Thus one of the benefits of an assessment program is that it fosters–indeed, necessitates–reflection on program and course goals. In his influential study of scholarship for the Carnegie Foundation, Ernest Boyer identified reflective critique as one of the key principles underlying assessment practices of students, faculty, programs, and higher education [Glassick, 1997]. Indeed, unless linked to an effective process of reflection, assessment can easily become what many faculty fear: a waste of time and effort.

But what if the faculty want more heat and the students need more humidity? How do we find that out if we only measure the temperature? It is not uncommon for mathematics faculty to measure success in terms of the number of majors or the number of graduates who go to graduate school, while students, parents, and administrators may look more to the support mathematics provides for other subjects such as business and engineering. To ensure that goals are appropriate and that faculty expectations match those of others with stakes in the outcome, the assessment cycle must from the beginning involve many constituencies in helping set goals. Principle two: *Assessment must be an open process.*

Almost certainly, a goal-setting process that involves diverse constituencies will yield different and sometimes incompatible goals. It is important to recognize the value of this variety and not expect (much less force) too much uniformity. The individual backgrounds and needs of students make it clear that uniform objectives are not an important goal of mathematics assessment programs. Indeed, consensus does not necessarily yield strength if it masks important diversity of goals.

The purpose of assessment is to gather evidence in order to make improvements. If the temperature is too low, the thermostat turns on the heat. The attribution of cause (lack of heat) from evidence (low temperature) is one of the most important and most vexing aspects of assessment. Perhaps the cause of the drop in temperature is an open window or door, not lack of heat from the furnace. Perhaps the cause of students' inability to apply calculus in their economics courses is that they don't recognize it when the setting has changed, not that they have forgotten the repertoire of algorithms. The effectiveness of actions taken in response to evidence depends on the validity of inferences drawn about causes of observed effects. Yet in assessment, as in other events, the more distant the effect, the more difficult the attribution. Thus principle three: *Assessment must promote valid inferences.*

Compared to assessing the quality of education, taking the temperature of a home is trivial. Even though temperature does vary slightly from floor to ceiling and feels lower in moving air, it is fundamentally easy to measure. Temperature is one-dimensional, it changes slowly, and common measuring instruments are relatively accurate. None of this is true of mathematics. Mathematical performance is highly multidimensional and varies enormously from one context to another. Known means of measuring mathematical performance are relatively crude–either simple but misleading, or insightful but forbiddingly complex.

Objective tests, the favorite of politicians and parents, atomize knowledge and ignore the

interrelatedness of concepts. Few questions on such tests address higher level thinking and contextual problem solving–the ostensible goals of education. Although authentic assessments that replicate real challenges are widely used to assess performance in music, athletics, and drama, they are rarely used to assess mathematics performance. To be sure, performance testing is expensive. But the deeper reason such tests are used less for formal assessment in mathematics is that they are perceived to be less objective and more subject to manipulation.

The quality of evidence in an assessment process is of fundamental importance to its value and credibility. The integrity of assessment data must be commensurate with the possible consequences of their use. For example, informal comments from students at the end of each class may help an instructor refine the next class, but such comments have no place in an evaluation process for tenure or promotion. Similarly, standardized diagnostic tests are helpful to advise students about appropriate courses, but are inappropriate if used to block access to career programs. There are very few generalizations about assessment that hold up under virtually all conditions but this fourth principle is one of them: *Assessment that matters should always employ multiple measures of performance.*

Mathematics assessment is of no value if it does not measure appropriate goals–the mathematics that is important for today and tomorrow [MSEB, 1993; NCTM 1995]. It needs to penetrate the common facade of thoughtless mastery and inert ideas. Rhetorical skill with borrowed ideas is not evidence of understanding, nor is facility with symbolic manipulation evidence of useful performance [Wiggins, 1989]. Assessment instruments in mathematics need to measure all standards, including those that call for higher order skills and contextual problem solving. Thus the content principle: *Assessment should measure what is worth learning, not just what is easy to measure.*

The goal of mathematics education is not to equip all students with identical mathematical tool kits but to amplify the multiplicity of student interests and forms of mathematical talent. As mathematical ability is diverse, so must be mathematics instruction and assessment. Any assessment must pass muster in terms of its impact on various subpopulations–not only for ethnic groups, women, and social classes, but also for students of different ages, aspirations (science, education, business) and educational backgrounds (recent or remote, weak or strong).

As the continuing national debate about the role of the SAT exam illustrates, the impact of high stakes assessments is a continuing source of deep anxiety and anger over issues of fairness and appropriate use. Exams whose items are psychometrically unbiased can nevertheless result in unbalanced impact because of the context in which they are given (e.g., to students of uneven preparation) or the way they are used (e.g., to award admissions or scholarships). Inappropriate use can and does amplify bias arising from other sources. Thus a final principle, perhaps the most important of all, echoing recommendations put forward by both the Mathematical Sciences Education Board [1993] and the National Council of Teachers of Mathematics [1995]: *Assessment should support every student's opportunity to learn important mathematics.*

Implementations of Assessment

In earlier times, mathematics assessment meant mostly written examinations–often just multiple choice tests. It still means just that for high-stakes school mathematics assessment (e.g., NAEP, SAT), although the public focus on standardized exams is much less visible (but not entirely absent) in higher education. A plethora of other methods, well illustrated in this volume, enhance the options for assessment of students and programs at the postsecondary level:

- *Capstone courses* that tie together different parts of mathematics;
- *Comprehensive exams* that examine advanced parts of a student's major;
- *Core exams* that cover what all mathematics majors have in common;

- *Diagnostics exams* that help identify students' strengths and weaknesses;
- *External examiners* who offer independent assessments of student work;
- *Employer advisors* to ensure compatibility of courses with business needs;
- *Feedback* from graduates concerning the benefits of their major program;
- *Focus groups* that help faculty identify patterns in student reactions;
- *Group projects* that engage student teams in complex tasks;
- *Individual projects* which lead to written papers or oral presentations;
- *Interviews* with students to elicit their beliefs, understandings, and concerns;
- *Journals* that reveal students reactions to their mathematics studies;
- *Oral examinations* in which faculty can probe students' understanding;
- *Performance tasks* that require students to use mathematics in context;
- *Portfolios* in which students present examples of their best work;
- *Research projects* in which students employ methods from different courses;
- *Samples* of student work performed as part of regular course assignments;
- *Senior seminars* in which students take turns presenting advanced topics;
- *Senior theses* in which students prepare a substantial written paper in their major;
- *Surveys* of seniors to reveal how they feel about their studies;
- *Visiting committees* to periodically assess program strengths and weaknesses.

These multitude means of assessment provide options for many purposes–from student placement and grading to course revisions and program review. Tests and evaluations are central to instruction and inevitably shine a spotlight (or cast a shadow) on students' work. Broader assessments provide summative judgments about a students' major and about departmental (or institutional) effectiveness. Since assessments are often preludes to decisions, they not only monitor standards, but also set them.

Yet for many reasons, assessment systems often distort the reality they claim to reflect. Institutional policies and funding patterns often reward delaying tactics (e.g., by supporting late summative evaluation in preference to timely formative evaluation) or encourage a facade of accountability (e.g., by delegating assessment to individuals who bear no responsibility for instruction). Moreover, instructors or project directors often unwittingly disguise advocacy as assessment by slanting the selection of evaluation criteria. Even external evaluators often succumb to promotional pressure to produce overly favorable evaluations.

Other traps arise when the means of assessment do not reflect the intended ends. Follow-up ratings (e.g., course evaluations) measure primarily student satisfaction, not course effectiveness; statements of needs (from employers or client departments) measure primarily what people think they need, not what they really need; written examinations reveal primarily what students can do with well-posed problems, not whether they can use mathematics in external contexts. More than almost anything else a mathematician engages in, assessment provides virtually unlimited opportunities for meaningless numbers, self-delusion, and unsubstantiated inferences. Several reports [e.g., Stenmark, 1991; Stevens, 1993; Shoenfeld, 1997] offer informative maps for navigating these uncharted waters.

Assessment is sometimes said to be a search for footprints, for identifying marks that remain visible for some time [Frechtling, 1995]. Like detectives seeking evidence, assessors attempt to determine where evidence can be found, what marks were made, who made them, and how they were made. Impressions can be of varying depths, more or less visible, more or less lasting. They depend greatly on qualities of the surfaces on which they fall. Do these surfaces accept and preserve footprints? Few surfaces are as pristine as fresh sand at the beach; most real surfaces are scuffed and trammeled. Real programs rarely leave marks as distinguishing or as lasting as a fossil footprint.

Nevertheless, the metaphor of footprints is helpful in understanding the complexity of assessing program impact. What are the footprints left by calculus? They include cognitive and attitudinal changes in students enrolled in the class, but also impressions and reputations passed on to

roommates, friends, and parents. They also include changes in faculty attitudes about student learning and in the attitudes of client disciplines towards mathematics requirements [Tucker & Leitzel, 1995]. But how much of the calculus footprint is still visible two or three years later when a student enrolls in an economics or business course? How much, if any, of a student's analytic ability on the law boards can be traced to his or her calculus experience? How do students' experiences in calculus affect the interests or enthusiasm of younger students who are a year or two behind? The search for calculus footprints can range far and wide, and need not be limited to course grades or final exams.

In education as in industry, assessment is an essential tool for improving quality. The lesson learned by assessment pioneers and reflected in the activities described in this volume is that assessment must be broad, flexible, diverse, and suited to the task. Those responsible for assessment (faculty, department chairs, deans, and provosts) need to constantly keep several questions in the forefront of their analysis:

- Are the goals clear and is the assessment focused on these goals?
- Who has a voice in setting goals and in determining the nature of the assessment?
- Do the faculty ground assessment in relevant research from the professional literature?
- Have all outcomes been identified–including those that are indirect?
- Are the means of assessment likely to identify unintended outcomes?
- Is the mathematics assessed important for the students in the program?
- In what contexts and for which students is the program particularly effective?
- Does the assessment program support development of faculty leadership?
- How are the results of the assessment used for improving education?

Readers of this volume will find within its pages dozens of examples of assessment activities that work for particular purposes and in particular contexts. These examples can enrich the process of thoughtful, goal-oriented planning that is so important for effective assessment. No single system can fit all circumstances; each must be constructed to fit the unique goals and needs of particular programs. But all can be judged by the same criteria: an open process, beginning with goals, that measures and enhances students' mathematical performance; that draws valid inferences from multiple instruments; and that is used to improve instruction for all students.

References

American Mathematical Association of Two-Year Colleges. *Crossroads in Mathematics: Standards for Introductory College Mathematics Before Calculus.* Memphis, TN: American Mathematical Association of Two-Year Colleges, 1995.

Committee on the Undergraduate Program in Mathematics (CUPM). "Assessment of Student Learning for Improving the Undergraduate Major in Mathematics." *Focus: The Newsletter of the Mathematical Association of America,* 15:3 (June 1995) 24-28.

Douglas, Ronald G. (Editor). *Toward a Lean and Lively Calculus* Washington DC: Mathematical Association of America, 1986.

Glassick, Charles E., et. al., *Scholarship Assessed: Evaluation of the Professoriate.* Carnegie Foundation for the Advancement of Teaching. San Francisco, CA: Jossey-Bass, 1997.

Hoaglin, David C. & Moore, David S. (Editors). *Perspectives on Contemporary Statistics.* Washington, DC: Mathematical Association of America, 1992.

Frechtling, Joy A. *Footprints: Strategies for Non-Traditional Program Evaluation.* Washington, DC: National Science Foundation, 1995.

Joint Policy Board for Mathematics. *Recognition and Rewards in the Mathematical Sciences.* Providence, RI: American Mathematical Society, 1994.

Loftsgaarden, Don O, Rung, Donald C., & Watkins, Ann E. *Statistical Abstract of Undergraduate Programs in the Mathematical Sciences in the United States: Fall 1995 CBMS Survey.* Washington, DC: The Mathematical Association of America, 1997.

Madison, Bernard. "Assessment of Undergraduate Mathematics." In *Heeding the Call for Change: Suggestions for Curricular Action,* Lynn A. Steen, editor. Washington, DC: Mathematical Association of America, 1992, pp. 137-149.

Mathematical Sciences Education Board. *Moving Beyond Myths: Revitalizing Undergraduate Mathematics.* Washington, DC: National Research Council, 1991

Mathematical Sciences Education Board. *Measuring What Counts: A Conceptual Guide for Mathematics Assessment.* Washington, DC: National Research Council, 1993.

National Council of Teachers of Mathematics. *Assessment Standards for School Mathematics.* Reston, VA: National Council of Teachers of Mathematics, 1995.

National Council of Teachers of Mathematics. *Curriculum and Evaluation Standards for School Mathematics.* Reston, VA: National Council of Teachers of Mathematics, 1989.

Roberts, A. Wayne (Editor). *Calculus: The Dynamics of Change.* Washington, DC: Mathematical Association of America, 1996.

Schoenfeld, Alan. *Student Assessment in Calculus.* Washington, DC: Mathematical Association of America, 1997.

Steen, Lynn Arthur (Editor). *Calculus for a New Century: A Pump, Not a Filter.* Washington, DC: Mathematical Association of America, 1988.

Steen, Lynn Arthur (Editor). *Reshaping College Mathematics.* Washington, DC: Mathematical Association of America, 1989.

Stenmark, Jean K. (Editor). *Mathematics Assessment: Myths, Models, Good Questions, and Practical Suggestions.* Reston, VA: National Council of Teachers of Mathematics, 1991.

Stevens, Floraline, et al. *User-Friendly Handbook for Project Evaluation.* Washington, DC: National Science Foundation, 1993.

Tucker, Alan C. and Leitzel, James R. C. *Assessing Calculus Reform Efforts.* Washington DC: Mathematical Association of America, 1995.

Wiggins, Grant. "A True Test: Toward More Authentic and Equitable Assessment." *Phi Delta Kappan,* May 1989, 703-713.

Wiggins, Grant. "The Truth May Make You Free, but the Test May Keep You Imprisoned: Toward Assessment Worthy of the Liberal Arts." *The AAHE Assessment Forum,* 1990, 17-31. (Reprinted in *Heeding the Call for Change: Suggestions for Curricular Action,* Lynn A. Steen, editor. Washington, DC: Mathematical Association of America, 1992, pp. 150-162.)

| **Home Page** | **Selected Papers** | **Top of Page** |

Last Update: August 18, 1999. *Contact:* Lynn A. Steen
URL: http://www.stolaf.edu/people/steen/Papers/assessment.html
Copyright © 1999. Disclaimer